chapter 11

GENES, PROTEINS, AND EVOLUTION

" . . . we must remember that heredity, development, and evolution are essentially epigenetic and not preformistic. We do not inherit from our ancestors, close or remote, separate characters, functional or vestigial. What we do inherit is, instead, genes which determine the pattern of developmental processes. . . . "

T. DOBZHANSKY, *Evolution, Genetics, and Man.*

As the genes of a species are modified and reshuffled, occasional organisms will appear within a restricted population having phenotypic characteristics that enable them to explore desirable ecological niches which were unattainable by their predecessors. The individual changes are generally quite small. Many generations must come and go, during which forays into formerly forbidden territory by this developing branch of the population become more fre-

quent and are of longer duration as the result of further reorganization of the gene pool by random mutation and natural selection. In time the summation of these changes results in a new species, fully at home in its new environment and sufficiently different in physiology from its distant ancestors that cross-fertilization is no longer possible.

We have discussed, in several earlier chapters, the techniques used by the geneticist for the analysis and description of limited portions of such a chain of events. As long as crosses can be made between different family lines, phenotypic changes can generally be related to specific genes and the spread of these genes through a population can be fairly accurately mapped. Thus, the basic assumptions of evolutionary theory may be directly tested, and with some precision, when the segment of time under consideration is small, and we are able to describe the process in terms of changes in *genotype*. When we deal with evolution on a larger scale, however, the tools of the geneticist are no longer applicable. The evolutionist must now rely on the study of relative morphology and ecology as deduced from the fossil record, or on the comparative anatomy and physiology of living representatives of surviving species.

The principal aim of this book has been to examine the basic principles underlying another possible method for the study of evolution. This method is based on the hypothesis that the individual proteins which characterize a particular species are unique reflections of the genes which control their synthesis. The examination of the chemistry of a series of homologous proteins is, of course, a purely phenotypic approach to the problem. Nevertheless, the evidence available to us, even at this early date, suggests that the structure of proteins may be a relatively direct expression of gene structure and that comparative protein chemistry may furnish a qualitative view of genotypic differences and similarities. If we accept the general hypothesis, we are led to infer, for example, that the "insulin-determining" genes of the pig and the sperm whale are identical, like the insulins whose structures they determine. Indeed, should several genes be concerned with the synthesis of insulin, the same would also be true for these.

Another interesting potentiality of comparative protein chemistry is that it might permit us to determine whether the same phenotypic characteristic, shown by two completely unrelated organisms, is attributable to *analogous* or to *homologous* genes. For example, both bacteriophage T2 and chicken's eggs contain proteins that have lysozyme activity. The genetic material of both coliphage and chickens must be said to contain information that can direct the formation of

proteins with this function. Is it possible that these two organisms contain nearly identical (that is, homologous) stretches of genetic material, or are the genes for lysozyme synthesis, and the lysozymes themselves, entirely different? This would appear to be the sort of question that might be attacked directly by the comparative study of protein structure. As we have already seen, in connection with cytochrome c, ribonuclease, hemoglobin, and other proteins, there is excellent evidence which indicates that many homologous genes *do* appear to have survived happily through long periods of time, some well exceeding the span of the fossil record.

A Biochemical Approach to the Species Problem

The paleontologist, in estimating the rates and directions of evolution, must depend almost entirely on morphological evidence. Even with this relatively crude sort of yardstick, he can begin to distinguish patterns of change such as we discussed in Chapter 1, in connection with the characteristics of tooth structure in the evolving horses. He is limited, however, to the *results* of evolution and can never hope to elucidate the underlying physiological changes that participate to produce new phyla.

Although most of the ancient species disappeared, representatives of almost all the phyla escaped extinction by adapting to their new environments, thus perpetuating large parts of heredity. We have available to us, then, a contemporary sample of the life of the past from which we should be able to deduce a great deal about the factors that were decisive in phylogenesis long ago. The study of "biochemical evolution" has already been of considerable value in the establishment of biological interrelationships. For example, the occurrence of melanocyte-stimulating, oxytocic, and vasopressor hormonal activity in extracts of the neural gland if tunicates furnishes strong evidence in support of the assignment of this subphyllum, the Urochordata, to the direct pathway between the invertebrates (which lack MSH activity) and vertebrates. The presence of both arginine phosphate (an invertebrate phosphagen) and creatine phosphate (the typically vertebrate phosphagen) in tunicates adds additional support to this assignment.

We shall not attempt to discuss here the numerous contributions of this sort that biochemistry has made to evolutionary theory. The reader will find this material summarized in a number of comprehensive essays and books.[1, 2, 3] Our present concern is primarily with

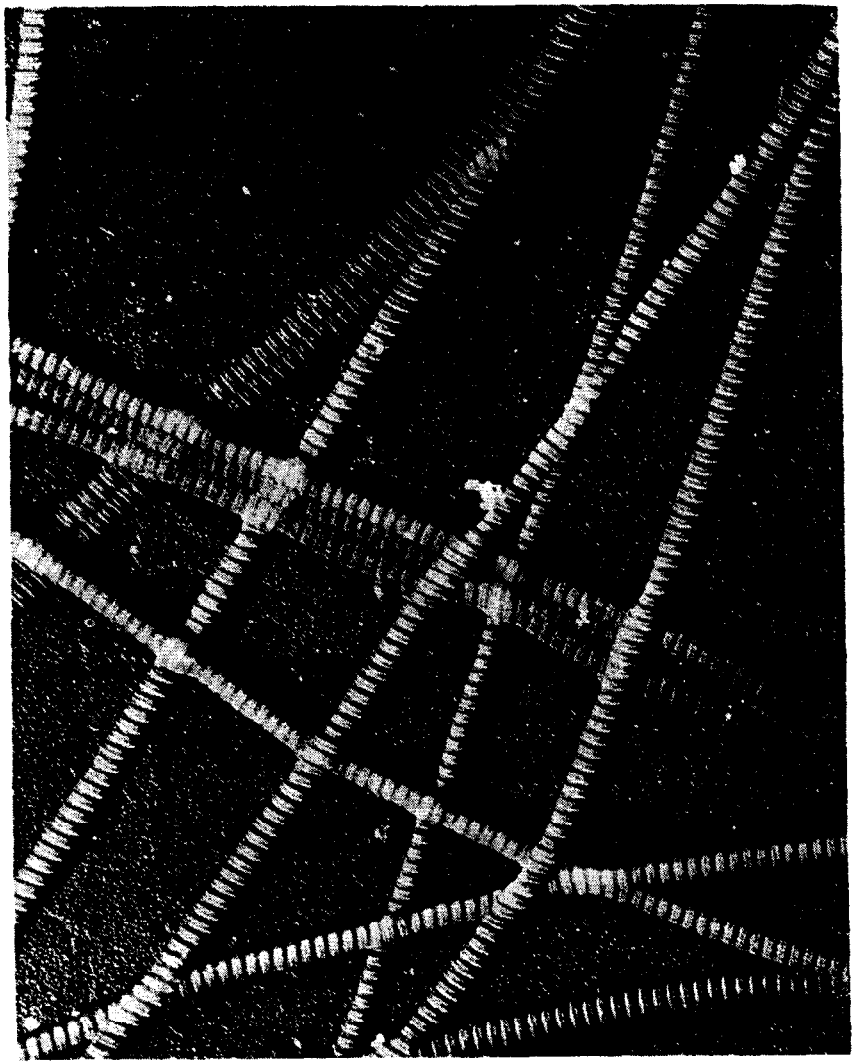

Figure 96. An electron photomicrograph of collagen fibrils from bovine skin. Magnification × 42,000. Obtained through the kindness of Dr. Jerome Gross, Massachusetts General Hospital, Harvard University Medical School.

the biochemical changes in protein molecules that are much nearer, in metabolic terms, to the genes themselves than are such products of enzymatic action as the phosphagens, or the eye pigments of *Drosophila*. As Wald has put it "It is a truism in biochemistry that each species of animal and plant possesses specifically different proteins." The full understanding of speciation must, almost certainly, be sought in the structure of proteins. To expand this point a bit, let us consider two elegant examples of speciation that are demonstrably related to changes in protein structure.

The protein collagen is largely responsible for the physical properties of such structural tissues as skin and cartilage. When collagen fibrils (Figure 96) are exposed to heat they change markedly in internal structure and yield the molecular form known as gelatin. Now recent studies by X-ray crystallography have shown that the collagen molecule is very likely composed of three strands of polypeptide, cross-linked through a system of hydrogen bonds of considerable strength.[4] The amino acid sequence, Gly.Pro.Hypro, appears fairly frequently along the chains, and the hydroxyl groups on the hydroxyproline residues are presumably major contributors to the hydrogen bond network. When collagen is heated in solution, the hydrogen bonds become ruptured at a critical temperature, known as the "shrinkage temperature," and the organized structure is quickly disoriented to form the more globular and amorphous gelatin structure. Although the exact mechanism of this rearrangement is not known, it is possible, on the basis of the results of current work on the properties of synthetic polyproline and of mixed polymers of proline and glycine, that the shrinkage may be associated with a *cis-trans* isomerization at proline-proline or proline-hydroxyproline bonds,[5] in conjunction with ordinary "entropic" denaturation.

On the basis of these chemical and physical observations we might suppose that collagen molecules, suited to either cold or warm habitats, could be devised by nature through the introduction or deletion of hydroxyproline residues. Animals living in climates tending to be very warm would do well to utilize collagens with high shrinkage temperatures, and those living in cold climates could do with considerably fewer sites for cross-linkage and with lower shrinkage temperatures.

The studies of K. H. Gustavson and of T. Takahashi on the collagens of fishes suggest that this is precisely the mechanism which has been employed.[6] The shrinkage temperatures of cold-water fishes are always lower than those of warm-water fishes, and an amazingly linear relationship exists between shrinkage temperature and the con-

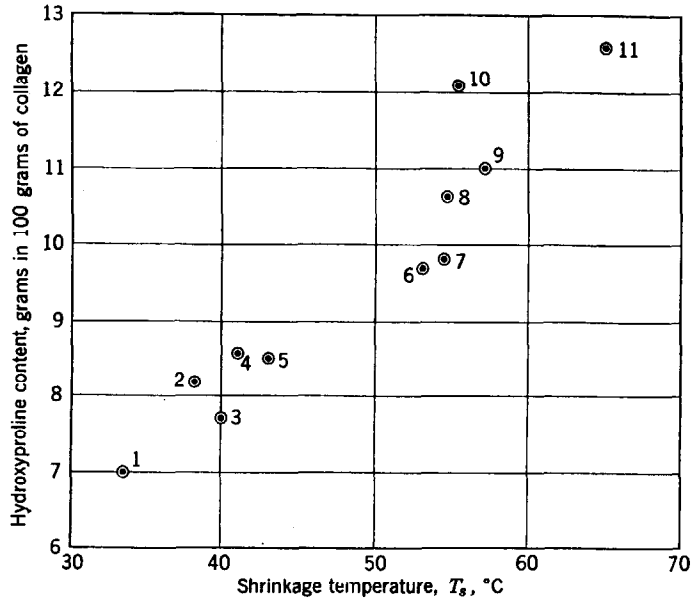

Figure 97. The relationship between the hydroxyproline contents of the collagens of various fishes and their "shrinkage temperature."[6]

tent of hydroxyproline (Figure 97), although the vertebrate collagens are otherwise extremely similar in composition. It is a provocative fact that collagen shrinkage temperatures seem to fall about 15 or 20° above the highest temperatures likely to be encountered by a species, as though this margin of safety were adequate in the ordinary course of climatic events.

The relationship between the visual pigments of marine fishes and the depths of their habitats is another dramatic example of adaptation through modification of protein structure. Denton and Warren,[7] Munz,[8] Wald and his colleagues,[9] and many other investigators have studied the chemical structure and the spectral properties of a variety of fish rhodopsins. Rhodopsin, composed of a vitamin A derivative complexed with a protein, opsin, constitutes the light-sensitive element of the retinal rods. The vitamin A-like prosthetic group, retinene, responsible for light absorption, has been found to be identical in all the species listed in Table 17. Since opsins do not themselves absorb light in the spectral interval between 480 and 503 mμ, the shifts in the position of the absorption maxima shown in Figure 98 must be attributed to the effects of the opsins on the

TABLE 17
Spectral Properties of Rhodopsins from Various Fishes

Species	Summer Range of Depth, fathoms	λ_{max}, mμ	E_{540}/E_{max}
Summer flounder (*Paralichthys dentatus* Linnaeus)	2–10	503	0.695
Scup (*Stenotomus versicolor* Mitchell)	1–20	498	0.586
Butterfish (*Poronotus triacanthus* Peck)	1–30	499	0.610
Barracuda (*Sphyraena borealis* DeKay)	1–10	498	0.575
Cod (*Gadus callarias* Linnaeus)	5–75	496	0.530
Cusk (*Brosme brosme* Müller)	10–100	494	0.455
Lancet-fish (*Alepisarus ferox* Lowe)	>200	480	0.250

spectral properties of retinene. The mechanism by which conjugation with opsin can induce a change in the spectral properties of retinene is quite obscure. We have previously discussed a related instance of a spectral shift, where the absorption characteristics of the

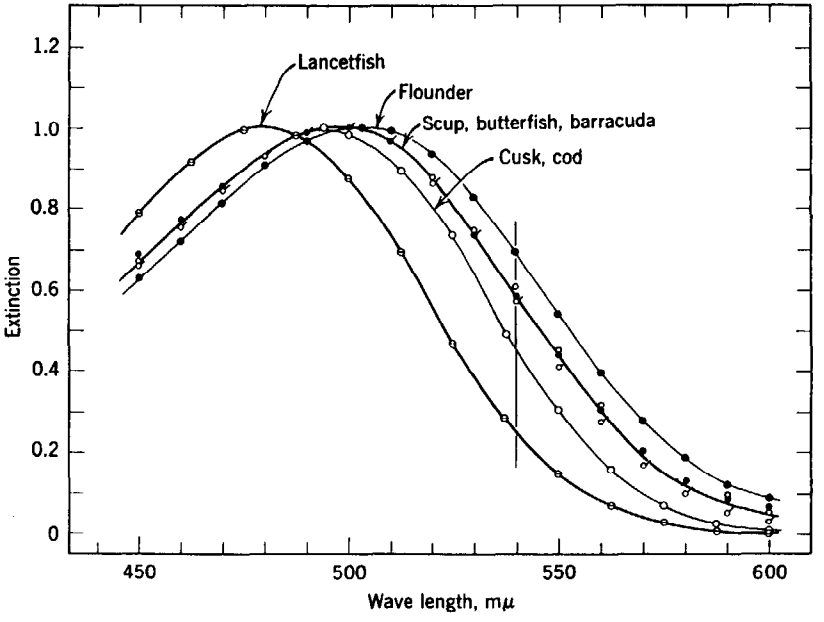

Figure 98. Absorption spectra of rhodopsins of marine fishes in 2 per cent aqueous digitonin solution. The maximum absorption (λ_{max}) shifts toward shorter wavelengths in rough correlation with the depth of habitat. See Table 17.

tyrosine residue are modified by hydrogen bonding of the hydroxyl group. The shift in this case was small, of the order of a few millimicrons. In the rhodopsin absorption system maxima differ by as much as 23 $m\mu$ as we move from the summer flounder to the lancet fish. This large shift implies a major change in the nature of the interaction between protein and prosthetic group.

The biological observation that makes all this of special interest is the fact that a correlation is observed between the mean depth of habitat of the various species of fishes and the spectral properties of their visual pigments. The correlation is not at all exact and, as the data in Table 17 show, a wide spread of λ_{max} exists at all depths. Nevertheless, the information available is sufficient to form the basis of a strong hypothesis. Over twenty years ago G. L. Clarke observed that the increasing blueness of light with depth in the ocean raises "the question of the possibility of a shift in the sensitivity of the eye of a deep-water fish toward the blue end of the spectrum." This possibility is realized in the spectroscopic observations just listed, and it is now possible to apply the techniques of protein chemistry to the elucidation of the details of this fascinating chapter in biochemical ecology. The study of the structural modifications in opsin which have taken place during the evolution of the fishes will be especially interesting since, as we have seen for the insulins and cytochromes c, large spans of evolutionary time may pass without too extensive a change in a particular protein molecule. The changes in the opsin molecule may be so cleverly contrived and so incisive that extensive alterations in sequence and folding have been unnecessary. On the other hand, if alterations *have* been extensive, we shall be required to rationalize a very complex set of interactions between protein and prosthetic group. Both alternatives are intriguing, to say the least, and the study of the chemistry of the opsins should make a most valuable contribution to the understanding of evolution at the molecular level.

The Rate of Evolution

As G. G. Simpson has pointed out, the question "How fast has evolution occurred?" is meaningless without the addition of the qualifications, "the evolution of what organisms, of which of their structures, and at what time in their history." The opossum, for example, has changed relatively little in the past 80,000,000 years, whereas the

evolution of the horses during the past 60,000,000 years has involved at least eight distinct genera.

Just as the anatomical organization of some organisms has changed much more rapidly than that of others, it seems likely that we shall find a large spread in the rates at which specific protein molecules have been modified during evolution. Although our basis for discussion of this point is still very thin, it is already evident that some proteins have undergone far greater structural change than others over an equivalent period. Compare, for example, the somatotropins of the sperm whale and the sheep with the insulins of these same species. Although the changes in insulin structure have been restricted to very minor modifications in a limited part of one polypeptide chain, the somatotropins are quite markedly modified in molecular weight, in cystine content, and in the number of polypeptide chains. Equally striking differences in degrees of modification exist between numerous others of the examples discussed in Chapter 7.

How are we to plan our experimental approach in attempting to establish some chemical coherence in the tremendous puzzle of speciation? We must, it would seem to me, begin with the basic assumption that the phenotypic character of a species is primarily determined by its unique spectrum of proteins. We may then proceed to a study of the extent to which each of the individual proteins within any spectrum may be modified without loss of biological function. As we already know, the degree of "violability" of different proteins may vary enormously as judged from the results of *in vitro* studies on denaturation and chemical modification in relation to function. Even here, however, many of the observed differences in sensitivity may be overemphasized and may depend on the choice of methods used for modification. Even though two proteins may be very similar in regard to the proportion of their total structure that is essential for function, one set of reagents may attack critical parts of one and not seriously alter the other. Amino groups, for example, may be acetylated with essentially no effect in pepsin, but at least some of these same groups appear to be critical for the activity of lysozyme. A proper comparison of two biologically active proteins thus must depend on the use of a wide variety of inactivating reagents, and ultimately on the deliberate degradative sort of study that aims to reduce proteins to their minimum, functionally adequate size.

Since, however, proteins *can* be modified without loss of function, it seems certain that the permissible degree of modification, in terms of fractions of their total structure, will vary somewhat from molecu-

lar species to species. It does not seem too farfetched to think of the proteins of a given organism as being subdivisible into those that have structures quite closely tailored to an essential functional requirement, those that are designed with only moderate "efficiency" or whose function is relatively dispensable, and those that are intermediate. Once again, illustrations come to mind. Several individuals exhibiting only slight clinical abnormality have been shown to be *completely* devoid of serum albumin. These individuals, able to lead a normal existence, are living evidence for the dispensability of this protein under the ecological circumstances peculiar to humans. On the other hand, no one will question the inability of most species to survive in the absence of cytochrome c or of the enzymes necessary for oxidative phosphorylation.

If we accept these subdivisions of the protein spectrum, we may "express" a species in terms of a hierarchy of protein structures ranging in violability from none to very much. The further evolution of this species, involving the usual mutation and natural selection, would then be reflected in a change in its proteins, one end of the spectrum remaining relatively fixed while the other may change considerably. Thus the cytochrome c molecule, which we might think of as a relatively "primitive" protein, and indispensable for most life, would be stubbornly perpetuated in the evolving phyla with minimal change, whereas the structure of the serum albumins might fluctuate with the shifting parameters of natural selection. From time to time, entirely new protein structures might arise, as in the dramatic appearance of insulin and of other hormones at the point in evolution when the protovertebrates and vertebrates appeared. The molecular basis for such "explosive" appearance of new protein entities is, of course, completely obscure. Only a thorough understanding of the processes of protein biosynthesis and of genetic information transfer will enable us to choose between such alternatives as the *de novo* creation of a whole new gene as opposed to the fortuitous reshuffling of already available genetic units.

We may safely predict that the patterns of change observed in the protein spectrum by future biochemists will not always be smooth and tidy. The criteria of natural selection will differ greatly from species to species, from environment to environment, and from period to period, and the survival value of gene mutations in a population, and of their images in the phenotype, will be quite varied.

A large number of important aspects of evolution have been omitted from this book. Some of these omissions may be attributed to type-

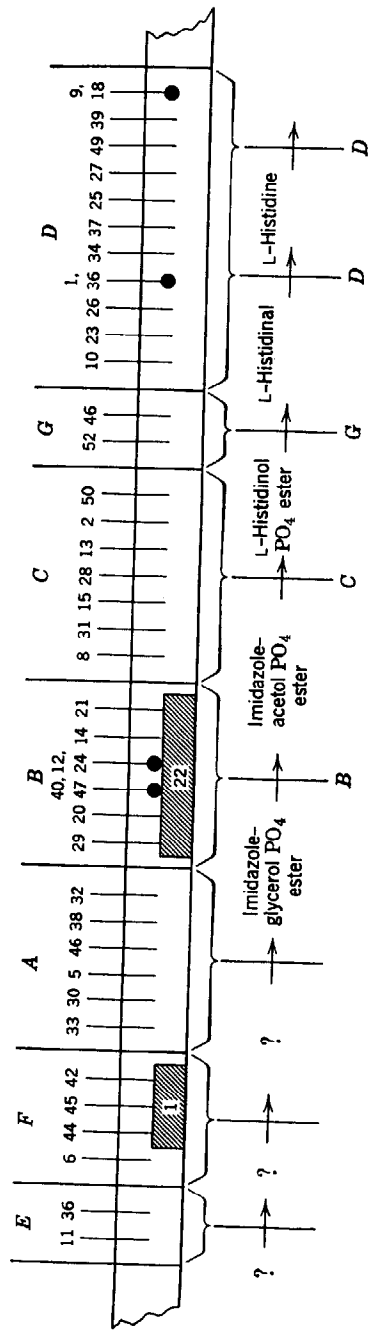

Figure 99. Linkage map of the region of the chromosome of *Salmonella typhimurium* which controls the stepwise synthesis of histidine. The enzymes which catalyze the series of reactions shown at the bottom of the figure are presumably synthesized under the control of various "cistrons" labeled *E, F, A, B, C, G.* and *D.* The order of the individual mutations within any cistron is only tentative. These mutations were mapped by use of the "transduction" technique described in Chapter 4. These fascinating results suggest that in *Salmonella* the cistrons corresponding to the biosynthetic enzymes are arranged in the same order as the biosynthetic steps themselves. For details consult the review by Philip E. Hartman in *The Chemical Basis of Heredity,* Johns Hopkins Press, 1957.

writer fatigue. Most of them, however, have been purposely omitted because of the lack of adequate factual material for discussion, and this book is already well supplied with speculation. We might, for example, have taken up the question of the spatial organization of genes in relation to function. The recent studies on the mapping of genes related to histidine biosynthesis in *Salmonella* (Figure 99) by Hartmann, Demerec, and others indicate that the various cistrons associated with the series of intermediate enzymes occur in the same region of the genetic strand and that these genetic determinants are arranged on the gene map *in the same order as the reaction sequence itself*. A schematic representation of this linkage map is given in Figure 99. The evolutionary implication that linked biochemical steps have been added, successively, in sequence along the chromosome is a very exciting one but is clearly not general. In *Neurospora*, for example, genetic loci for closely related enzymatic steps are scattered at random throughout the chromosomal apparatus.

We might, also, have spent some time on the question of cytoplasmic heredity, which we know to be of importance in many biological systems. Here again the scarcity of published information is a limiting factor. The study of inheritance of traits in a non-Mendelian fashion is likely to be difficult and confusing, and the genetic or biochemical study of such traits might receive disproportionately little attention. As Nanney has recently suggested, "It is perhaps only natural that investigations of 'messy' characteristics are discontinued before publication and that investigators move on to traits more readily analyzed." The omnivorous reader will find Nanney's review[10] of the subject of cytoplasmic heredity in *The Chemical Basis of Heredity* excellent reading.

The list of omissions can be extended. The chemistry of RNA and its genetic properties, the rearrangements of genes within the chromosome and the phenotypic consequences of such rearrangements, the problem of polyploidy, the interactions of nonallelic genes—many of these might, even now, be discussed with some intelligence in biochemical terms.

The relationships between genotype and phenotype will, predictably, become a major preoccupation of more and more "pure" and medical scientists during the coming years. This book has grown out of my own attempts to arrive at some sort of appreciation of the potentialities of chemical genetics and the evolutionary approach. It will have been well worth the effort if it can help to stimulate the growing interest in evolution as the central theme in the life sciences.

REFERENCES

1. J. B. S. Haldane, *The Biochemistry of Genetics*, Allen & Unwin, London, second impression 1956.
2. M. Florkin, *Biochemical Evolution*, Academic Press, New York, 1949.
3. G. Wald, in *Modern Trends in Physiology and Biochemistry*, Academic Press, New York, 1952.
4. A. Rich and F. H. C. Crick, *Nature*, **176**, 915 (1955).
5. W. F. Harrington, *Nature*, **181**, 997 (1958).
6. K. H. Gustavson, *The Chemistry and Reactivity of Collagen*, Academic Press, New York, 1956. The analytical studies of Takahashi are discussed on pages 224–227.
7. E. J. Denton and F. J. Warren, *Nature*, **178**, 1059 (1956).
8. F. W. Munz, *Science*, **125**, 1142 (1957).
9. G. Wald, P. K. Brown and P. S. Brown, *Nature*, **180**, 969 (1957).
10. D. L. Nanney, *The Chemical Basis of Heredity* (W. D. McElroy and B. Glass, editors), Johns Hopkins Press, Baltimore, 1957.